
**Independent Peer Review Report on the STAR Panel for
Sablefish and Dover Sole, held 25–29 July 2011, in Newport,
Oregon**

Sven Kupschus

Prepared for

Center for Independent Experts (CIE)

Centre for Fisheries and Aquaculture Science
Lowestoft Laboratory
Pakefield Road
Lowestoft
Suffolk NR33 0HT
United Kingdom

Phone +44 1502 524454
e-mail sven.kupschus@cefas.co.uk
www.cefas.co.uk



Contents

Executive summary	3
1. Background	5
2. Review activities	5
3. Findings	6
4. Conclusions / Recommendations	8
Appendix 1. Bibliography of materials provided for review	14
Appendix 2: Statement of work	16
Appendix 3: STAR participants	26

Independent Peer Review Report on the STAR Panel for Sablefish and Dover Sole, held from 25th-29th July 2011, in Newport, Oregon.

Executive Summary

Activities

The 2011 draft assessment reports on sablefish and Dover sole and the supporting documentation were provided according to the scheduled timescale. All documentation was reviewed thoroughly ahead of the panel meeting. The review meeting was conducted through a series of presentations by the stock assessment team (STAT) followed by questions from the Panel. The panel sought to understand the linkages in the stock dynamics implied by model implementation and to ascertain their appropriateness given the understanding of biology, ecology and fisheries for the two stocks. This, in conjunction with the examination of the model diagnostics, allowed for panel requests to the STAT to be sufficiently focused to be effective in addressing the questions of appropriateness of the model through examination of its sensitivity to assumptions and describing the major sources of uncertainty.

The outcome of the week's work was that both draft assessments represented the best scientific information that could be extracted from available information. In addition, the process highlighted some weaknesses in both assessments which should be investigated in future assessments, but which could not be addressed by this panel. However, these were not thought to be so severe as to render them inappropriate for management advice, particularly in the light of the low exploitation level of either stock.

Main findings for Dover sole:

The choices of data sources were well described and justified, with at least an equal amount of effort being devoted to an extensive investigation into possible model parameterisation and their consequences conducted by the author of the assessment report prior to the meeting. This essential preparation considerably aided the panel's discussion and its ability to interpret the model results, particularly with respect to uncertainty.

In contrast, previous assessments for this stock also implemented in-stock synthesis (SS3 v3.21), the 2011 draft assessment base case presented a less complex modelling approach, to increase parsimony. The model was examined in detail and several alternative models were investigated on the basis of the contributions of various data sources to the likelihood profile, residual diagnostics and differences in the selectivity estimated. Changes to the base model included altering the choice of selectivity function for specific surveys and fisheries, and an in-depth investigation of the sensitivity of the assessment to estimating natural mortality, M , and how this could result in an apparent link between the estimate of male M and current female spawning-stock biomass (SSB).

The results illustrated some weaknesses in model implementation, but also suggested that the effects of this on management would be minor. Therefore, the draft base case developed by the STAT was accepted as the best available scientific information and appropriate for management in the current framework. I personally have confidence in this verdict and feel that the assessment provides an improved understanding of stock dynamics over previous assessments, and highlights some of the vagaries of the SS3 implementation. Given the counterintuitive implementation of stock dynamics in the model, it seems unlikely that the estimates of uncertainty are very accurate, but having understood the process behind this and the relatively low level of exploitation of the stock, I was able to gauge this problem as a relatively minor issue for the assessment.

One unfortunate effect of the dynamics, however, is that both virgin SSB and current SSB are very uncertain, which could cause overly precautionary management if the alternate states of nature provided by the panel are implemented in the P* approach. The STAR panel is instructed to evaluate whether alternative states of nature are on the scale of the uncertainty of current SSB based on the major axes of uncertainty in other parameters. This represents a wide range in this assessment. However, actual management is based on the ratio of current to virgin SSB (spawning potential ration, SPR), which is much less affected by the problem in the assessment because both SSB estimates vary in unison, not independently, as applied by the directive.

Main findings for sablefish:

The focus of the sablefish assessment was to try to rationalize the complexity of previous assessments. While there can be little doubt that the population dynamics of sablefish on the west coast, not to mention the development of the fisheries themselves, are significantly more complex than allowed for by this model, it is clear that data availability is also limited, especially historically. The STAT carried out an extensive preliminary analysis in developing the draft model and provided much of the knowledge gained during this process. Providing such a good summary allowed the panel to understand better and more quickly the choices and dynamics of the draft base model.

The model was characterized by very wide confidence limits in virgin SSB, but less uncertainty about current SSB. This might have compromised management of the stock particularly as the current SPR is below the current management target of SPR45%. However, it is clear from the assessment and data that a number of recent below-average recruitments have contributed to this decline, not simply overfishing, with total catches less than the annual catch limit (ACL) and significantly less than the overfishing limit (OFL).

With the exception of steepness, all other model parameters (including growth) were estimated freely in the model. This greatly reduced the need for a sensitivity analysis because the uncertainty is formally incorporated if the deviates are randomly distributed. However, model output illustrated consistent patterns in the residuals, which suggested that the model was somehow constrained in terms of matching the true stock dynamics. Much of the

investigation and requests to the STAT focused on elucidating the nature of this constraint and/or conflict between data sources, to determine if and by how much this would impede the provision of sound management advice. It was clear that some quite dramatic changes to model configuration, although not resolving the issues regarding the residual patterns, did not greatly alter the management advice, suggesting that the model is sufficiently robust and the patterns in residuals sufficiently small not to make the model sensitive to alternate specifications. Although I feel confident that the management advice implied by the model will not in the short term lead to undesirable levels of exploitation, I understand less of how the assessment arrives at its conclusions than, for example, the Dover sole assessment, so I feel less able to adjudicate its utility for long-term management. Personally, therefore, I believe that this model should be treated as a benchmark assessment for future assessments.

A number of other sensitivity runs had been conducted, but the panel conclusion was that most of the uncertainty was already included in the model and that the understanding of M represented the major source of uncertainty. Therefore, the draft base case developed by the STAT was accepted as the best available scientific information and appropriate for management.

Important recommendations:

A number of formal recommendations were advanced by the panel to improve future assessments, and these are specifically included in the panel report. The most important with respect to the Dover sole assessment was that the interdependence between male and female selectivity in SS3 needed to be resolved because it leads to undesirable constraints in the assessment. That made it difficult to investigate other aspects of the assessment, such as the apparent disparity in the sex ratio from different data sources. For sablefish, further research is needed to address the questions of if and why recent strong year classes have been missed. However, the way forward in terms of specific recommendations, other than to take a general look at age information and sampling strategy is not clear. An important consideration for a much wider range of assessments to consider for future assessments is the question of whether the confidence limits based on a maximum likelihood estimate (MLE) approach are appropriate, because there was some indication, at least for Dover sole, that the full Markov chain Monte Carlo (MCMC) probability distributions were significantly more skewed than the MLE approximation. Interestingly, it is the sablefish assessment which would lend itself best to this examination.

The above recommendations are in keeping with the recommendations in the STAR Panel Summary Report, but represent only a subset of those presented in the panel report that I consider most important for improvement of the assessments.

Independent Peer Review Report on the STAR Panel for Sablefish and Dover Sole, held from 25th-29th July 2011, in Newport, Oregon.

1. BACKGROUND

In accordance with the Statement of Work (SOW: Appendix 2), I was contracted to participate as a CIE independent review panellist for the 2011 STAR Panel 4 for Dover sole and sablefish. This document represents my own findings and interpretation of the information provided, and is based on the panel meeting and discussions. However, some of the thoughts and conclusions were formulated in the process of writing this report, so may not be identical to those provided in the final official panel report.

2. REVIEW ACTIVITIES

The 2011 STAR4 Panel for Dover sole and sablefish was held at the Northwest Fisheries Science Center (NWFSC), Newport, Oregon, from the 25 to 29 July 2011. The Terms of Reference for the STAR Panel are given in the SOW (Appendix 2: Annex 2).

Panel members, members of the STAT, the SSC and other participants in the review are listed in Appendix I. The meeting was open to the public, and was attended by several observers. For each stock, the results of the assessments were presented to the STAR panel and other attendees, and the input data, assessment approach, results and utility of the findings for management were evaluated through open discussion. In the course of discussion, the Review Panel requested that additional analyses and evaluations be carried out by the STAT. These requests were documented and presented to the STAT, who undertook the analyses requested and provided appropriate feedback to the Panel.

TOR 1

Prior to the Review Workshop, I was provided with draft stock assessment reports for sablefish and Dover sole, and an update / correction to the sablefish model (Appendix 1). These were made available according to the agreed timescale via an ftp site. The documents were thoroughly reviewed ahead of the review meeting in order to gain a full understanding of the rationale for the approach, and of the input data and assumptions used for the stock assessment. Small alterations to both assessments were necessary following the release of the documentation of the site, because of minor corrections to the model (sablefish) and corrections to the catch data (Dover sole), although these did not alter either the perception of stock status or the assessment diagnostics discernibly.

TOR 2

The presentations on data provided by the STAT were very informative and, on the whole, the data were found to be of a quality appropriate for the methods

applied to them. There were some points which should be considered in the future.

- a) For both assessments the catch data had to be compiled from a variety of data bases by the STAT. Ideally this would be handled by the relevant data collection experts and provided in a uniform database format. Here, a small revision was required to the Dover sole assessment due a late discovery of an error.
- b) For sablefish, the sampling levels by state over time were examined. This indicated at time substantial changes in the contribution of samples from different states which has the potential to impact that assessment given the spatial distributions of different fisheries and the uncertainty regarding spatial differences in growth. A more centrally organised sampling program may help, though it appears that the situation has been improving recently.
- c) Discussions during the Dover sole sensitivity analysis indicated that the sex ratio estimation between the length and the age information provided by the same survey was inconsistent. This was the case for two surveys (the ages from one of these were not used in the assessment). The sampling protocol should be examined to eliminate this as the cause.

TOR 3, 4 and 6

Although it was apparent at this panel that SS3 still has some minor difficulties (separating male and female selectivities, not able to accurately reproduce assessments carried out in previous versions due to different implementation) the general approach and the methodology is a sound one. The model has a near infinite number of settings making it easy to get lost in the detail and hence overparameterise the assessments. I appreciated the STAT's approach to these two stocks starting with a simple model and implementing changes only when there was independent evidence that such alterations could be supported. This enabled the models to converge even while estimating M despite low information content on this parameter. M proved to be major axes of uncertainty in both assessments, and elegantly allowed the use of the MLE confidence limits to estimate the alternate states of nature for sablefish. This was not the case for Dover sole, because of the selectivity problem in SS3; however a sensitivity analysis over various levels of male M proved robust and appropriate method to assess the risk.

TOR 5

Both draft assessments were deemed to provide the best scientific information available for their respective stocks.

The panel reports for each stock were completed in rough draft by the end of the review meeting, but minor editing and reviewing for accuracy and consistency between stocks continued until 8 August.

TOR 7

The panel was an efficient process, with the STAT panel interactions being cooperative and helpful. Generally, the discussions resulted in everyone understanding better the how and why. Differences in opinion were rare and, if present, confined to minor issues such as the order of prioritisation of the research recommendations or the potential cause / speculation as to the origin of a particular symptom. In all cases a compromise could be reached.

Comments on the STAR process

I found the STAR process not only an effective and thorough one, but also an interesting and enjoyable experience. Much of this, however, was attributable to the energy and enthusiasm shown by the STAT in preparing for the panel, and their knowledge regarding the respective biology and fisheries. Usually at these types of meetings, an indication of the success of the assessment process is the number of industry people attending and the time they spend there. Few industry members, as at this meeting, can either mean the process has failed and industry has lost faith in the system, or as here that they believe the process to be in good hands and carried out appropriately. For reviewers, it is helpful to have more industry representatives present to understand better the processes behind the fisheries and to attain an impression of the historical development of the stock. In this case, the STAT had already spent a considerable amount of time communicating with the industry, so this was clearly much less of an issue and undoubtedly a key reason why attendance was low despite the considerable value of these fisheries to the west coast community.

I appreciated the rigorous process of submitting research requests and the reasoning behind such requests, and I think that the process worked very well here. However, from the perspective of a reviewer, the tendency is to ask more questions than can be answered in the available time, and that could potentially lead to unacceptably high workloads for the assessment staff at the STAT. Here, focusing the questions and hence reducing the number of requests in conjunction with the technical expertise and work carried out by the STAT prior to the meeting resulted in a successful outcome. However, had the STAT been less thorough or less responsive to the requests, or had there been a need to make large-scale alterations to the assessment implementation, such success would have been much less likely. I think this is worth bearing in mind with respect to the panel process implemented by the STAR.

3. FINDINGS

Dover sole

Introduction

Choices of data sources were described and justified in the draft assessment document. Presentation of this information at the meeting brought the actual

information content of each data source much more to life, which helped the panel very quickly to understand the modelling process. An extensive investigation into possible model parameterisation and their consequences, conducted by the author of the assessment report prior to the meeting, also considerably aided the panel discussion and led to the development of effective requests as to how to interpret the model results particularly with respect to uncertainty.

In contrast to previous assessments for this stock also implemented in SS3, the 2011 draft assessment base case presented a less complex modelling approach in order to increase parsimony in the model. One of the major changes to previous assessments was the fact that all catches (all taken by trawl fisheries) were attributed to one of three fleets based on the landings of each, with selectivity estimated for three time-blocks (1910-1980, 1981-1995 and 1996-2010) compared with the random walk selectivity applied previously to two fleets.

Available survey data used in the assessment were AFSC triennial (split into two periods), AFSC slope (1998-2002), and NWFSC slope and shelf-slope (2003- present). The survey data were pre-processed using GLMM models externally to the assessment model, to account mainly for spatial differences in the NWFSC survey extent, but also to account for differences in the efficiencies of boats/skippers, while maintaining the uncertainty structure appropriately in the assessment. The process used the delta-gamma approach, but given the relatively small number of zero catches, a Poisson approach would have been more appropriate, even when the information is represented as biomass (see research recommendations). For the central tendency of the indices this is likely to have had very little impact given the virtually ubiquitous distribution of the species, so for the purpose of this assessment, the current information was deemed appropriate and an improvement over previous practice (2005 assessment) of using solely the slope part of the survey now that more information is available on the wider population.

Investigations

Two areas of the assessment were conducted in more detail at the STAR panel meeting with respect to this assessment. The first concerned the use and appropriateness of the modelling with respect to the information contained in the NWFSC shelf slope survey (NWFSC combo), the other regarding the more general implementation in SS3 of the selectivity parameterisation and its consequences.

Selectivity information provided by the AFSC and NWFSC slope surveys was modelled as a cubic spline smoother, resulting in a bi-modal selectivity curve by length for both females and males. On closer examination, the second mode for males was determined to be in excess of the lengths recorded for males in the survey and was based on the interdependence of selectivity of females as implemented in SS3. Considerable discussion developed on the merits of the spline function. In this case, the model determines the bi-modality to be an artefact of sampling, whereas there is some information from previous

investigations of the survey that this may represent a real feature in the population, so that the model should not be allowed to adapt this level of complexity in order to more-accurately reflect the population dynamics in the stock. Several of those present were convinced that a less than completely mixed component of the stock exists in the southern region, corresponding to the large females of intermediate age discussed previously.

The panel concluded that to retain parsimony in the model until such time that such a discrete segment of the population could be identified independently and their contribution to the spawning stock confirmed, it would be more appropriate to accept the potential overparametrization in q as a likely bias in the assessment rather than dealing with the uncertainty of a discrete subpopulation for which it is not possible to assume the same biological or selectivity parameters and hence significant decreases in model parsimony (underestimation of true uncertainty) if the stock component was evaluated separately.

Another, and potentially more serious, concern with the assessment was whether the dynamics of the model implied uncertainty surrounding the value of M . It became apparent that current female SSB estimates depended on the estimation of male M , for which no biologically plausible reasoning existed. Closer examination of model properties identified the SS3 implementation of the male and female selectivities for commercial information as the cause. Commercial male selectivity was modelled asymptotically, and female selectivity using a double-normal approach linked through a common offset parameter. The idea behind this is that selectivity is essentially a length-based, gear-interaction term. However, when, as in this case, selectivity is a multi-process function made up of spatial segregation by gender in addition to the gear process, the dependence between the gender-specific selectivities is no longer appropriate. However, in the current implementation of SS3, it is not possible to decouple the selectivities because one is modelled as an offset of one to the other. Switching the gender to which the offset is applied is not recommended as a solution, because maximum male selectivity is greater than maximum female selectivity, and rescaling selectivities may produce gradient problems. Future implementations of SS should allow independent modelling of the selectivities, at the very least, to investigate the appropriateness the purely gear-based consideration implemented in the current version.

Given the process error, estimates of uncertainty output by the new base model are unlikely to be very accurate, and this would warrant a sensitivity analysis to examine the effect of the selectivity offset, but which could not be conducted here because of the inability to implement other states of nature in SS3.

Interestingly, one potential consequence of the linkage between selectivities is the scale of the current biomass estimate. One of the concerns of the panel was that the catchability (q) of the NWFSC combo survey is low given the spatial extent of the survey and the area-based approach used to calculate the index information. In the model, survey information is treated as a relative index of abundance, so that the low values of q estimated by the model result in considerably larger estimates of biomass than suggested by the survey alone.

At least part of the reason for this has to be the dependence of selectivity. The fishery catches significantly larger males than would be expected given the survey information, resulting in residuals in catch-at-age information, given the current modelling approach. However, there are insufficient older females in the catch information to decrease the estimates of q because the sex ratio in the survey information is significantly skewed, with more females in the raw survey age data and greater numbers of males in the length information.

One attempt to resolve the conflict using age-specific mortality (Lorenzen M) yielded a worse fit to the data for both length and age information. The hope was that a value of M inversely related to size could resolve some of the contrasts on sex ratio at the larger size/older age, given the observed sexual dimorphism. However, the small differences in size-at-age at older ages meant that little of the uncertainty could be resolved on this basis.

The process error in the new base model does raise concerns regarding its appropriateness for management. However, the consequence of this error is the difficulty in estimating gender specific M , which already presented the major uncertainty within the assessment model and it was extensively investigated in the sensitivity analysis with respect to determining the alternate states of nature. This work, conducted during the panel and as part of the TOR for determining the columns of the forecast table, determined that given the exploitation history of the stock, and the overall uncertainty regarding the exploitation of the stock in the near future, the risk of overexploitation was minor. However, use of the assessment output to develop appropriate estimates of F_{MSY} do need to be conducted with significantly greater caution, because the values are likely highly dependent on an appropriate estimate of M . Consequently, the panel accepted the assessment as appropriate for management in the current circumstances, but recommended a number of possible improvements to the model, particularly the decoupling for gender-specific selectivity curves (see also research recommendations).

The methodology used for incorporating uncertainty into future catch streams as part of the TOR was developed around the model uncertainty in M , because this was the major uncertainty in the model, along with that caused by process error. The SSC requested the development of different states of nature on the basis of the interquartile range in uncertainty in the current biomass. This information is presented in the STAR panel report. However, the panel needs to point out that this may be inappropriate in this and potentially other cases, because the aim is to set catches for future years that, according to the current proxies of MSY (maximum sustainable yield) exploitation (depletion) do not depend only on the uncertainty in current levels of SSB , but also on the estimation of virgin SSB . When, as in this case, the major axis of uncertainty is M , the two estimates vary in unison, so that the uncertainty in depletion is smaller on a relative scale than the uncertainty in current SSB . The difference in the range of uncertainty is likely to be small to negligible for stocks that are heavily exploited, but for marginally exploited stocks with little information on M , this may result in management measures that will lead to exploitation significantly below F_{MSY} . Such measures may also lead to an autocorrelative process where information

density on M decreases further, leading to a spiral of underexploitation through ever-increasing uncertainty.

Despite some minor problems, the draft base case developed by the STAT was accepted as the best available scientific information available, and appropriate for management in the current framework. I am personally confident in this verdict and feel that the assessment provides an improved understanding of the stock dynamics over previous assessments, while still highlighting some of the vagaries of SS3 implementation. The uncertainty in the management parameters is captured well in the assessment, although I feel that the STAR directives for assessing alternative states of nature on the scale of the uncertainty of current SSB is inappropriate for this stock. Uncertainty in current SSB varies in Unison with uncertainty in virgin SSB and that relationship is near linear (due to the low exploitation of the stock. Consequently, the spawning potential ratio (here the F_{MSY} proxy and hence relevant management measure) is far less sensitive to the uncertainty in M than current SSB. The danger with using current SSB as the measure of uncertainty is that management action based on precautionary principles may reduce catches further which will lead to a spiral of increasingly sparse information due to fewer samples, higher uncertainty and future decreases in ACL for a stock that by all measures appears to be under exploited for market reasons at this point.

In order to make improvements to the current assessment model, it must first be possible to resolve the interdependence of male and female selectivity in SS3. Without this model revision, it is not possible to examine any further the remaining sources of sex-ratio conflict (contrasting with the length and age data of a single survey, and the relative contribution at age to the fishery between sexes), so at this point it is difficult to provide exact guidance as to what needs to be investigated further.

Sablefish

Introduction

The focus of this assessment was to attempt to rationalize the complexity of previous assessments. While there can be little doubt that the population dynamics of sablefish on the west coast, not to mention the development of the fisheries, are significantly more complex than allowed for by the model, it is clear that data availability is also limited, especially historically.

Modelling of commercial selectivities using flexible spline functions can be seen as a concession to overall model simplification. Much time can be spent trying to interpret the often bi-modal and frequently asymmetrical nature of the estimates, but there is usually little evidence that these models are interpretable in this way. Here, the interpretation was more that these shapes, although complex, would be more likely to reflect the average dynamics in the populations, rather than trying to account specifically for each characteristic of the data.

Survey data were assumed to be more consistent than commercial information, so warranted more-restrictive selectivity modelling. The use of the double-normal function does allow for dome-shaped selection in survey data, thought to be appropriate because even at a depth of 1200 m not the entire population of sablefish is covered, with an unknown proportion found at greater depth.

The resultant base model was characterized by very wide confidence limits in virgin SSB, but less uncertainty about current SSB. This might have compromised management of the stock, particularly as the current spawning potential ratio is below the current management target of SPR45%. However, it is clear from the assessment and data that a number of recent below-average recruitments have contributed to this decline rather than overfishing, with total catch less than ACL and significantly less than OFL.

With the exception of steepness, all other model parameters (including growth) were estimated freely in the model using iterative re-weighting, rather than making *a priori* decisions on the value of specific data sources. This greatly reduced the need for a sensitivity analysis because the uncertainty is formally incorporated, but it assumes that deviates are randomly distributed, whereas the model output illustrated some consistent patterns in the residuals that suggested that the model was somehow constrained in terms of matching the true stock dynamics.

Investigations

Much of the investigation and requests to the STAT focused on elucidating the nature of this constraint or conflict between data sources, and to determine its severity with respect to impeding the provision of sound management advice. Specifically, the model appeared to be underestimating the contribution of recent good year classes, which was of particular concern because the 2008 year class appeared to be abundant and about to enter the fishery. Also, all three sets of commercial length information contained more large fish than expected by the model after 1999, and all three sets of commercial data indicated an overabundance of older fish that was consistent over a number of years during different periods.

Changes in the selectivity blocking and functions applied, changing the plus group age in the data, and considering age-variant values of M were unable to reduce the pattern in residuals significantly relative to the base model, nor could an investigation of temporal changes in sampling effort be linked with these patterns. However, the sometimes dramatic changes in model configuration made had relatively little impact on relative stock trajectories, suggesting that the model is reasonably robust. The investigations also confirmed that the model contained very little information on the absolute scale of SSB, with virgin SSB being particularly poorly understood.

Another aspect of the assessment that I still do not fully understand is the estimated decline in the biomass from 1980 to 1990, which appears not to be supported by any of the information. The NWFSC-combo survey does imply a decline, but more recently. Landings too declined rapidly during that period

owing to the exclusion of foreign fleets then, but declining landings are only representative of declining biomass at constant effort, for which there is no suitable information. Removing the older age information did not change the current picture much, although it did remove a number of good recruitments historically, suggesting that there were reasonable year-class signals in the age information at ages >15. Such trend information was not immediately apparent from the raw age data and may be the source of the implied decline in biomass.

The panel concluded that these patterns in the residuals were relatively small overall, and in many ways had to accept that the commendable effort to simplifying model structure was always likely to result in some systematic biases. However, it was deemed that this was still preferable to an overparametrized model that was likely to be just as biased, but would significantly underestimate the uncertainty in management output. In future assessments, my opinion is that it could be useful to devote time to investigating the origins of the cohort underestimates. A number of possible reasons could be:

- a) Although recruitment deviate CVs were set high, they are still expected to follow a lognormal trend, whereas they may be strongly over dispersed
- b) There was a suggestion that age-determination error may be involved
- c) Conflict in the length information with that in the age information due, for example, to changes in the length-at-age for different cohorts could result in a similar effect if modelled as a single growth curve for the population.

There was very little information in the data regarding steepness (fixed at $h = 0.6$). The information that may have been present was absorbed by the great variation in recruitment deviates implemented in the model, so the sensitivity of the assessment management output with respect to the assumption on steepness was small at reasonable levels of steepness ($h > 0.3$). However, the value of F_{MSY} is highly sensitive to this assumption and cannot be inferred on the basis of this assessment.

A number of other sensitivity runs had been conducted, but the panel conclusion was that most of the uncertainty was already included in the model and that M represented the main source of uncertainty.

Therefore, the draft base case developed by the STAT was accepted as the best available scientific information and appropriate for management. From a personal perspective, I am convinced that the current model will be sufficient to provide management advice in the short term, but there are a number of issues still not fully understood regarding why the model comes up with the results that it does. Investigations along these lines could prove fruitful in future.

Recommendations for future research and data collection

General (affecting more than one assessment)

1. Complete and review the Washington catch reconstruction and review the California and Oregon catch reconstructions. The accuracy and wide availability of consistent basic information is essential to the development of Pacific coast assessments. In addition to the raw data, the reliability and availability of more spatially disaggregated forms of the data should be investigated to determine whether they could be used to develop more spatially explicit models without sacrificing accuracy.
2. SS3 needs to be modified to disable the gender offset function even if this complicates convergence. The difficulties encountered in the Dover sole assessment and some other flatfish assessments with respect to the linkage between selectivities do require addressing in future. Although in many instances, size-based selectivity may be appropriate, when sexes separate spatially, there is a requirement for models to at least be able to investigate complete independence between genders. It is important that this be implemented in an updated version of SS3, if for no other reason than to be able to eliminate it as the cause of some of the sex-ratio conflicts in the assessment.
3. Develop guidelines for use of the Lorenzen model for age-dependent natural mortality. The panel investigated the use of age-specific natural mortality in both assessments presented during STAR 4. In each case, one of the reasons for exploring different mortality schedules was the difficulty in fitting the unbalanced abundance-at-age information (as seen through residuals to fits), either in the sex ratio at older ages (Dover sole) or the ratio of young to old fish (sablefish). The use of Lorenzen's M based on a decline in natural mortality by the inverse of the growth rate implies a link with predation. However, wider use and development of some guidance on the appropriateness of the implementation in other stock assessments should be investigated.
4. Include in future versions of stock synthesis the capability to explore alternative error-distribution assumptions for compositional data. Currently, the only available error distribution for age information is the multinomial probability function. It appears that this may have some impact with respect to underestimating strong year classes, and it would be desirable to explore the use of alternative error assumptions in order to analyse survey information, in particular where variance estimates in catches at age may be less than independent of abundance.
5. Conduct new studies of maturity at length and age, based on more comprehensive coastwide and depth-based sampling, and using histological techniques for determining maturity stage. Given that there is

uncertainty regarding the temporal stability of maturity schedules, there should be periodic monitoring to explore for changes in maturity.

6. Update the STAR Terms of Reference to ensure that assessment documents include standard plots (or tables) of likelihood profiles that include likelihood components by data source and fleet. Such plots are an important diagnostic tool for displaying tensions among data sources.

Specific to Dover sole

1. Age-determination bias must be investigated. This is important for Dover sole given the current base model difficulty in reconciling some tensions between different data sources on sex ratio at the oldest ages. In addition, the ability of the model to track cohorts accurately would be significantly disrupted if there were severe size-based bimodality in cohorts caused by vastly different times of settlement (Dover sole are thought to have a larval period of 6–18 months). Consequently, larval period should also be examined.
2. Sampling protocol and procedures for the NWFSC combo survey should be re-examined to ensure that samples are representative of the catches. Raw age and length information appeared to imply persistently different sex ratios when viewed in isolation. The concern here is that there is some unrepresentative sampling of the age distribution as ages are subsampled from length. The sampling procedure should be investigated more closely, and potentially improved.
3. Publish the report of the NMFS workshop. The conclusions of the NMFS workshop on developing priors on survey catchability were not available to the Panel. These should be made available and the information reconsidered specifically with respect to Dover sole, in an attempt to reconcile the relatively low catchability estimates for the surveys, particularly the NWFSC combo survey, which is thought to cover most of the stock distribution.
4. Continue to investigate spatial and temporal patterns in growth. Having simplified the model compared with previous assessments, especially with respect to uniform growth, it is important to continue investigating whether this is likely to introduce undesirable levels of bias into the assessment process as more information becomes available. Spatial information on the distribution by age/size of females, particularly in the southern part of the range and across the stratification boundaries of the survey as well as between stocks, should be the primary focus of this work.

Specific to sablefish

1. Further investigate potential inaccuracy in using maximum likelihood estimates and the normal distribution to approximate confidence limits for

estimates of spawning biomass. The current assessment's measures of uncertainty in spawning biomass are based on the assumption that the errors can be adequately approximated by normal distributions. The current model for sablefish is sufficiently simple that it may be feasible to conduct a full Bayesian analysis of uncertainty. There is concern that asymmetries in the error distributions, which the normal distribution cannot account for, may be creating a biased view of stock status.

2. A major uncertainty in the sablefish assessment relates to the maturity schedule and age determination. Better maturity and age-at-length data could reduce uncertainty and help resolve issues of cohort size.

Appendix 1. Bibliography of materials provided for the 2011 STAR Panel for Sablefish and Dover sole.

Barnes, T., Heppell, S., Mohn, R., Smith, S., and Thompson, G., 2005. Dover Sole
STAR Panel Meeting Report
June 20-24, 2005

Barnes, T., Heppell, S., Mohn, R., Smith, S., and Thompson, G., 2005.
Sablefish
STAR Panel Meeting Report
June 20-24, 2005

Dorn, M., Cordue, P., and Haist, V. 2007. Sablefish STAR Panel Report May 7-11, 2007

Hicks, A.C., and Wetzel, C., 2011. The Status of Dover Sole off the U.S. West Coast in 2011 (Draft--- Pre-STAR version)

Methot, R.D. 2011. User Manual for Stock Synthesis Model Version 3.21d (Updated May 8, 2011)

Sampson, D.B., 2005. The Status of Dover Sole off the U.S. West Coast in 2005.

Schirripa, M.J., 2007. Status of the Sablefish Resource off the Continental U.S. Pacific Coast in 2007

Schirripa, M.J., 2005. Status of the Sablefish Resource off the Continental U.S. Pacific Coasts in 2005

Stewart, I.J., Thorson, J.T., and Wetzel, C. 2011. Status of the U.S. Sablefish resource in 2011 (STAR panel draft – Revised executive summary only)

Appendix 2: Statement of Work for Dr. Sven Kupschus (CEFAS)

External Independent Peer Review by the Center for Independent Experts

Stock Assessment Review (STAR) Panel for Sablefish and Dover Sole

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: Benchmark assessments will be conducted for sablefish and dover sole. These are critical species to the slope fisheries and previous STAR panel reviews identified modeling issues to address. Additionally, the NWFSC shelf-slope survey data can only be fully included for the stock if full or benchmark assessments are conducted. Assessments for these two stocks will provide the basis for the management of the groundfish fisheries off the U.S. west coast, providing scientific basis for setting OFLs and ABCs as mandated by the Magnuson-Stevens Act. The technical review will take place during a formal, public, multiple-day meeting of fishery stock assessment experts. Participation of external, independent reviewer is an essential part of the review process. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Two CIE reviewers, one of which will participate during all STAR panels for consistency, shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in fish population dynamics, with experience in the integrated analysis modeling approach, using age-and size-structured models, use of MCMC to develop

confidence intervals, and use of Generalized Linear Models in stock assessment models. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Newport, Oregon during 25-29 July 2011.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: <http://deemedexports.noaa.gov/sponsor.html>).

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Documents to be provided to the CIE reviewers prior to the STAR Panel meeting include:

- The current draft stock assessment reports;
- Previous stock assessments and STAR Panel reports for widow rockfish;

- The Pacific Fishery Management Council's Scientific and Statistical Committee's Terms of Reference for Stock Assessments and STAR Panel Reviews;
- Stock Synthesis (SS) Documentation
- Additional supporting documents as available.
- An electronic copy of the data, the parameters, and the model used for the assessments (if requested by reviewer).

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

- 1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
- 2) Participate during the panel review meeting in Newport, Oregon during 25-29 July 2011.
- 3) In Newport, Oregon during 25-29 July 2011 as specified herein, and conducts an independent peer review in accordance with the ToRs (Annex 2).

- 4) No later than 12 August 2011, each CIE reviewer shall submit an independent peer review report addressed to the “Center for Independent Experts,” and sent to Mr. Manoj Shivilani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and to Dr. David Die, CIE Regional Coordinator, via email to ddie@rsmas.miami.edu.
- 5) Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

20 June 2011	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
11 July 2011	NMFS Project Contact sends the CIE Reviewers the pre-review documents
25-29 July 2011	Each reviewer participates and conducts an independent peer review during the panel review meeting
12 August 2011	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
26 August 2011	CIE submits CIE independent peer review reports to the COTR
31 August 2011	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:

- (1) each CIE report shall be completed with the format and content in accordance with **Annex 1**,
- (2) each CIE report shall address each ToR as specified in **Annex 2**,

(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COTR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivilani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

NMFS Project Contact:

Stacey Miller, NMFS Project Contact
NMFS Northwest Fisheries Science Center, 2032 SE OSU Drive, Newport OR 97365
Stacey.Miller@noaa.gov Phone: 206-437-5670

Michelle McClure
National Marine Fisheries Service, 2725 Montlake Blvd. E, Seattle WA 98112
Michelle.McClure@noaa.gov

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.
2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.
 - a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.
 - b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.
 - c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.
 - d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
 - e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.
3. The reviewer report shall include the following appendices:
 - Appendix 1: Bibliography of materials provided for review
 - Appendix 2: A copy of the CIE Statement of Work
 - Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Stock Assessment Review (STAR) Panel for Sablefish and Dover Sole

1. Become familiar with the draft stock assessment and background materials.
2. Comment on the quality of data used in the assessments including data collection and processing.
3. Evaluate and comment on analytic methodologies.
4. Evaluate model assumptions, estimates, and major sources of uncertainty and provide constructive suggestions for improvements if technical deficiencies or additional major sources of uncertainty are identified.
5. Determine whether the science reviewed is considered to be the best scientific information available.
6. Provide specific suggestions for future improvement in any relevant aspects of data collection and treatment, modelling approaches and technical issues.
7. Provide a brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations

Annex 3: Tentative Agenda

Note: Final Agenda to be provided two weeks prior to the meeting with draft assessments and background materials.

Stock Assessment Review (STAR) Panel for Sablefish and Dover Sole

Northwest Fisheries Science Center
Hatfield Marine Science Center,
2032 SE Oregon State University Drive,
Newport, Oregon, 97365

Monday, July 25, 2011

- 9:00 a.m. Welcome and Introductions
- 9:15 a.m. Review the Draft Agenda and Discussion of Meeting Format (Panel Chair)
 - Review Terms of Reference for Assessment and Review Panel
 - Assignment of reporting duties
 - Discuss and agree to format for the final assessment document
- 9:45 a.m. Stock Assessment Team (STAT-1) Presentation of Species 1 (Authors)
 - Overview of Data and Stock Synthesis Modeling
- 12:30 p.m. Lunch (On Your Own)
- 1:30 p.m. Q&A session with the STAT-1 & Panel discussion
- 3:30 p.m. Coffee Break
- 3:45 p.m. Panel develops request for additional model runs / analyses for STAT 1
- 4:30 p.m. Panel provides written requests for additional model runs / analyses to STAT 1
- 5:00 p.m. Adjourn for day.

Tuesday, July 26, 2011

- 9:00 a.m. Stock Assessment Team (STAT-2) Presentation of Species 2 (Authors)
 - Overview of Data and Stock Synthesis Modeling
- 12:00 p.m. Lunch (On Your Own)
- 1:30 p.m. Q&A session with the STAT-2 & Panel discussion
- 3:00 p.m. Coffee Break
- 3:15 p.m. Panel develops request for additional model runs / analyses for STAT 2
- 4:00 p.m. Panel provides written requests for additional model runs / analyses to STAT 2
- 4:30 p.m. Panel check in with STAT-1 if needed
- 5:00 p.m. Adjourn for day.

Stock Assessment Review (STAR) Panel for Sablefish and Dover Sole

Wednesday, July 27, 2011

- 9:00 a.m. STAT-1 Presentation of first set of model runs for Species 1
- Q&A session with the STAT-1 & Panel discussion
 - Panel develops written request for second round of model runs / analyses for STAT 1
- 12:00 p.m. Lunch (On Your Own)
- 1:30 p.m. STAT-2 Presentation of first set of model runs for Species 2
- Q&A session with the STAT-2 & Panel discussion
 - Panel develops written request for second round of model runs / analyses for STAT 2
- 3:30 p.m. Coffee Break
- 3:45 p.m. Continue Panel discussion with STAT-2
- 5:00 p.m. Adjourn for day.

Thursday, July 28, 2011

- 9:00 a.m. STAT-1 Presentation of Second Set of Model Runs for Species 1
- Q&A session with the STAT-1 & Panel discussion
 - Identification of preferred model and elements for the decision table.
 - Panel develops third list of model runs for decision table and begins drafting STAR report.
- 12:00 p.m. Lunch (On Your Own)
- 1:00 p.m. STAT-2 Presentation of Second Set of Model Runs for Species 2
- Q&A session with the STAT-2 & Panel discussion
 - Identification of preferred model and elements for the decision table.
 - Panel develops third list of model runs for decision table and begins drafting STAR report.
- 3:30 p.m. Coffee Break
- 3:45 p.m. Panel discussion or report drafting continues
- 5:00 p.m. Adjourn for day.

Friday, July 29, 2011

- 9:00 a.m. Consideration of remaining issues
- Review decision tables for Species 1 and Species 2
- 11:00 a.m. Panel agrees to process for completing final STAR report by Council's September meeting Briefing Book deadline
- 5:00 p.m. Review Panel Adjourn.

Appendix 3: Participants in the 2011 STAR Panel for Sablefish and Dover sole held from 25th-29th July 2011, NWFSC, Newport, Oregon, USA.

Technical Reviewers

Vidar Wespestad, Panel Chair, Scientific and Statistical Committee (SSC)
Kevin Stokes, Center for Independent Experts (CIE)
Sven Kupschus, Center for Independent Experts (CIE)
David Sampson, ODFW and Scientific and Statistical Committee (SSC)

Panel Advisors

John DeVore, Pacific Fishery Management Council (PFMC) Staff
Joanna Grebel PFMC Groundfish Management Team (GMT)
Gerry Richter, PFMC Groundfish Advisory Subpanel (GAP)

Stock Assessment (STAT) Teams

Ian Stewart, NMFS, Northwest Fisheries Science Center
Allan Hicks, NMFS, Northwest Fisheries Science Center
Chantel Wetzel, NMFS, Northwest Fisheries Science Center (*in absentia*)
James T. Thorson, University of Washington (*in absentia*)

Others Present:

Jim Hastie, NMFS NWFSC
Stacey Miller, NMFS NWFSC
Brad Pettinger, Oregon Trawl Commission
Tom Jagielo, Fishing Vessels' Owners Association representative
Lynn Mattes, Oregon Department of Fish and Wildlife (ODFW), GMT
Steve Bodnar, Coos Bay Trawlers Association
Craig Good, ODFW
Susan Hilber, ODFW
Patrick Milick, ODFW
Kelsey Adkisson, ODFW